nature research

Corresponding author(s): Charles G. Mullighan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	See below for software used. The software used for data collection also used for some level of data analysis.
Data analysis	In house software STRONGARM (RNA-seq mapping) CICERO (gene fusion detection) Bambino (SNV and Indel detection) SNPdetector Commercial software CGI Cancer Sequencing software (version 2.1) STAR (version 2.4.2a) BWA (version 0.7.12) Samtools (version 1.7) Picard (version 1.129) edgeR R package (v3.12.1) limma R package (v3.26.9) HTSeq (version 0.6.1p1) Affymetrix GeneChip Command Console Software v4.0.01567G Affymetrix Genotyping Console and the Birdseed v2

dChip (v1.0.0.1)
ZEN 2012 software (v1.1.1.0)
Enricher (see URLs)
R programming language (version 3.4.3)
ggplot2-2.2.1 package under R-3.3.2
SAS software (version 9.1.2)
Prism (GraphPad, version 7.0)
BD FACSDiva (version 8.0.1)
FlowJo (version 10)
CLC Main Workbench Version 7.7.2
Living Imaging v.4.4 software

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genomic data have been deposited in the European Genome-phenome Archive (EGA), 942 accession EGAS00001002537 and they may be explored interactively at the St. Jude PeCan Data Portal (https://pecan.stjude.cloud/proteinpaint/study/ael).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples were collected from different study groups according to the WHO 2008 and WHO 2016 criteria of classification after central review of pathology and immunophenotyping and 159 cases diagnosed with AEL were included in this analysis. No sample size calculation was performed. Given the rare nature of the diagnosis understudy, we collected as many samples as possible available to include in our analysis. This sample size is sufficient to provide an overview of common genomic profile of AEL and to allow further exploration into specific hypothesis based investigations.
Data exclusions	No data were excluded.
Replication	No replication of genomic analysis was performed in analysis
Randomization	Nothing to disclose. This is not an intervention study, so randomization is not applicable.
Blinding	Nothing to disclose. This is not an intervention study, so blinding is not applicable.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Flow analysis was performed using the following antibodies:
Biolegend: B220-BV605 (cat.# 103244, clone RA3-6B2, 1:20); Mac1 (CD11b) (cat.# 101237, clone M1/70, 1:20), Gr1-BV711 (cat.# 108443, clone RB6-8C5, 1:50), CD71-pe-Cy7 (cat.# 113812, clone RI7217, 1:50)
BD Biosciences: Gr1-PerCP-Cy™5.5 (cat.# 552093, clone RB6-8C5, 1:50), Ter119-V500 (cat.# 562120, clone TER-119, 1:50), CD3-APC (cat.# 553066, clone 145-2C11, 1:50), CD41-PE (cat.# 558040, clone MWReg30, 1:50), Mac1-Alexa700 (cat.# 557960, clone M1/70, 1:50), CD19-APC-Cy7 (cat.# 552854, clone 1D3, 1:50), CD34-BV421 (cat.# 562608, clone RAM34, 1:20), Ter119-BUV395 (cat.# 566206, clone TER-119, 1:20), CD44-APC (cat.# 559250, clone IM7, 1:50)
Termo Fisher Scientific: CD117-APC-eFluor780 (cat.# 47-1171-82, clone 2B8, 1:50).

Immunofluorescence was performed using the following antibodies:
Anti-6X His tag® antibody (abcam, cat.# ab9108, 1:200), Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 555 (Thermo Fisher Scientific, A-21428, 1:500)

For immunohistochemistry the following antibodies were used:
BD Biosciences: Rat Anti-Mouse CD45R/B220 (cat.# 553084, clone RA3-6B2, 1:1000), Rat Anti-Mouse CD34 (cat.# 553731, clone RAM34, 1:50), Rat Anti-Mouse CD43 (cat.# 552366, clone S7, 1:75), Rat Anti-Mouse CD45 (cat.# 553076, clone 30-F11, 1:500), Rat Anti-Mouse TER-119 (cat.# 550565, clone TER-119, 1:600)
Abcam: Rabbit anti-human GATA1 (cat.# ab131456, 1:4000), PAX5 antibody (cat.# ab109443, 1:125), Anti-RUNX1 / AML1 + RUNX3 + RUNX2 (cat.# ab92336, 1:1200)
Novus Biologicals: Integrin alpha 2b/CD41 Antibody (cat.# NBP1-84579, 1:1000),
Cloud clone: Polyclonal Antibody to Glycophorin A (cat.# PAB704Mu01, 1:50)
Dako: Anti Myeloperoxidase (MPO), (cat.# A0398, 1:1200)
SantaCruz: CD3-ε Antibody (cat.# sc1127, clone M-20, 1:1000)
Cell Signaling: TrkA Rabbit mAb (cat.# 2510, clone 12G8, 1:200)

Validation

All antibodies are validated for detecting mouse proteins by the manufacturer and confirmed for each specific application using cells of known origin and differentiation state and compared to isotype controls and cells that are known to express or lack the antigen. Fluorescence-labeled antibodies used for flow cytometric analysis were validated by the SJCRH Flow core facility.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

TF-1 and HEL cell lines were obtained from American Type Culture Collection (ATCC, Manassas, VA, USA). Cells were thawed and cultured according to ATCC's instructions.

Authentication

TF-1: Immunophenotype, karyotype, RNA-seq and whole exome sequencing, STR analysis. HEL: Immunophenotype, karyotype, RNA-seq and whole exome sequencing, STR analysis.

Mycoplasma contamination

All cell lines tested negative for micoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

Cell lines were not included in the database of commonly misidentified cell lines.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Mice included in this study included:
1) Female 8-week old wild-type C57BL/6 mice
2) Male and female 8-week old TP53 R172H knockin mice.

Wild animals

No wild animal were used in this study

Field-collected samples

No field-collected samples were used in this study

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study did not actually involve human subjects. We collected banked, de-identified tumor samples that were previously collected from patients with confirmed AEL. Those included: pediatric cases (n=35; age 0 to 20 years), young adults (n=8; age 21 to 39 years), adults (n=32; age 40 to 59 years), and older adults (n=84; age ≥ 60 years).
Recruitment	Patients were not recruited into the study. Only banked, de-identified patient sample were used for this study.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Samples included mouse primary cells (blood, cells from bone marrow, spleen and liver), lineage negative hematopoietic stem cells, TF-1 and HEL cell lines.
Instrument	Flow cytometric analysis and sorting was performed on an 18 color BD Biosciences Aria cell sorter
Software	BD FACSDiva was used to perform cell sorting and collect the data. FlowJo V10 was used to analyze the data.
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	Mouse blood and tissue samples for immunophenotype analysis were gated on live cells based upon SSC-A and FSC-A, then gated on singlets based upon FSC-A and SSC-W, then gated GFP+ or double GFP+/mCherry+ cells. Lineage negative hematopoietic stem cells: cells for sorting were gated on live cells based upon DAPI and FSC-A and sorted for GFP+ or double GFP+mCherry+.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.